

Approved for public release;
distribution is unlimited

C.1

Title:

Predictive Models for Transcriptional Enhancers

Author(s):

Goutam Gupta (B-8)*
Chang-Shung Tung (T-10)

Submitted to:

DOE Office of Scientific and Technical Information (OSTI)



Los Alamos

NATIONAL LABORATORY

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Predictive Models for Transcriptional Enhancers

Goutam Gupta* (B-8), Chang-Shung Tung (T-10)

Abstract

Our objective for the project is to develop a predictive method for studying and understanding the interactions between an enhancer element and its targeted binding protein. Based on homology modeling approach, we have developed a method for predicting structures of both the protein and DNA in the complex. Using homeodomain-DNA complex as our model system, we have shown that the prediction method is capable of generating structures that compare favorably with experimental results. The results from our study also provide insights to the sequence-specific recognition of the DNA by its targeted protein as well as the differences in binding due to site-directed mutations in the protein. Our prediction method is general, therefore, can be used for predicting structures of various protein-DNA complexes.

Background and Research Objectives

Being able to identify control elements such as transcriptional enhancers is an important step toward the understanding of gene expression. This process requires the knowledge of how these elements interact specifically with the corresponding proteins. The homeodomain is a small conserved protein motif that binds specifically to DNA and plays a central role in gene regulation. The homeodomain/DNA complex provide an ideal system for studying the interactions between gene control elements, such as enhancers, and their binding proteins. To accomplish this goal, we developed a method that is capable of predicting structures of both protein and DNA in a protein-DNA complex. Using this method, we studied the specific interactions between several homeodomain-DNA complexes.

Importance to LANL's Science and Technology Base and National R&D Needs

Homology modeling provides the crucial step in using the known structures derived from experiments and genome sequence data for predicting structures and functions of unknown proteins. We have furthered the homology modeling method to include the prediction of both the protein and DNA in a protein-DNA complex. Our prediction method is both novel and general and can be used for predicting structures of various protein-DNA complexes. This method should be very useful in the structural genome initiative.

*Principal Investigator email: gxg@lanl.gov

Scientific Approach and Accomplishments

Homology modeling is a method that predicts the structure of a protein (target) based on the known structure of a protein (template) that has homologous sequence [1]. This modeling approach has gained rapid popularity in recent years. Up-to-date, homology modeling remains the only method that can provide models with a root-mean-square error lower than 2 Å [2]. Following the basic concept of homology modeling, we have developed a method for predicting structures of both protein and DNA.

Using the engrailed homeodomain protein-DNA complex structure (1HDD in Protein Data Bank[3]) as the template (central molecule in Fig. 1a), we modeled the structures of the complex structures of both the antennapedia (left molecule in Fig. 1a) and the Mat α 2 (right molecule in Fig. 1a) homeodomains. The sequence identity between the engrailed and the antennapedia homeodomains is higher (50%) than that between the engrailed and the Mat α 2 homeodomains (30%). The root-mean-square errors for antennapedia homeodomain-DNA complex are 2.8 Å for protein (1.1 Å for C α) and 1.72 Å for DNA. The corresponding values are 3.0 Å (2.1 Å for C α) and 2.0 Å for Mat α 2 homeodomain-DNA complex. These results show that our modeled structures compare favorably with those determined experimentally.

Based on the structure of the engrailed homeodomain protein-DNA complex (1HDD in Protein Data Bank), we modeled the binding of the protein to different locations of the DNA. The molecule in the center of Fig. 1b shows the protein binds to the protein recognition sequence TTAATG (the corresponding base pairs are depicted as six horizontal lines). In our study, the DNA is moved 1 or 2 base pairs up and down with respect to the protein. With each of the different binding modes, the structure of the complex is energy minimized using AMBER[4]. The resultant structures are shown in Fig. 1b. The total number of hydrogen-bonds (H-bonds) between the protein and the DNA, as well as the H-bonds between the protein and the bases (listed in parenthesis), are listed under each of the complexes. While the total number of H-bonds between DNA backbone and protein remains unchanged, the number of H-bonds between DNA bases and protein is reduced by 2-3 when protein is moved from its optimum DNA binding position. This allows the sequence discrimination of the DNA binding for the protein.

In collaboration with Dr. Ma's group at the Children's Hospital Research Foundation (Cincinnati, Ohio), we studied the recognition codes of Bicoid homeodomain for different DNA sites. Bicoid homeodomain binds to both a TAATCC consensus site (A1) and a TAAGCT non-consensus site (X1). Both mutation and methylation interference studies indicate that arginine 54 (ARG-54) of the homeodomain plays the important role of base-specific recognition to the DNA. Using our homology modeling method, structures of Bicoid homeodomain binding to both A1 and X1 sites were developed. In our modeled structures, ARG-54 makes one hydrogen bond (h-bond) with the N7 atom of the adenine on the third position of the A1 site (see the left image of Fig. 1c). By swinging the residue upward about 1.5 Angstrom, ARG-54 makes two h-bonds (to the N7 atom of the adenine and to the O6 atom of the guanine 3' to the adenine) when the homeodomain binds to the X1 site (see the right image of Fig. 1c). Our results provide the explanation for the observation that the homeodomain mutant (arginine change to alanine at residue 54) binds to A1 site with a reduced binding affinity, while this specific mutation abolishes the binding of the homeodomain to the X1 site.

Publications

1. Tung, C.-S., "Structural Study of Homeodomain Protein-DNA Complexes using a Homology Modeling Approach." J. Biomol. Struct. Dyn. 17: 347-354, 1999
2. Dave, V., Zhao, C., Tung, C.-S., Ma, J. "Reprogrammable recognition codes of Bicoid homeodomain for different DNA sites." LA-UR-00-621, submitted to Genes and Development, 2000.

References

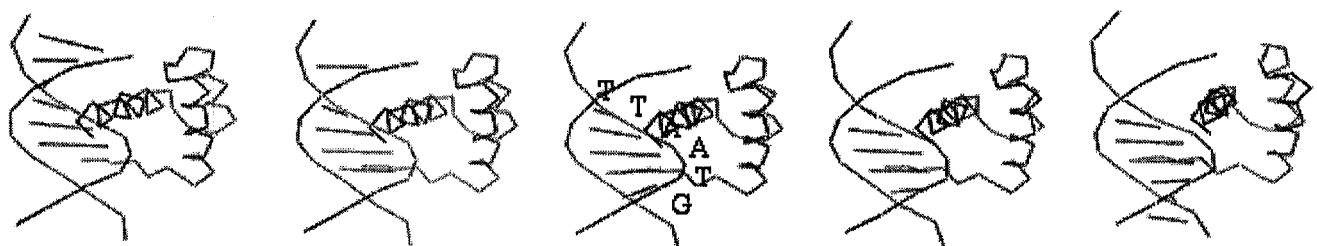
- [1] Sali A, Overington JP, Johnson MS, Blundell TL. "From comparisons of protein sequences and structures to protein modeling and design," TIBS 15: 235-240, 1990.
- [2] Sanchez R, Sali A. "Advances in comparative protein-structure modeling," Curr Opin Str Biol 7: 206-214, 1997.
- [3] Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. "Protein data bank: Computer-based archival file for macromolecular structures." J Mol Biol 112: 535-542, 1977.
- [4] Weiner SJ, Kollman PA, Nguyen DT, Case DA. "An all atom force field for simulations of proteins and nucleic acids." J Comp Chem 7: 230-252.

Figure

Figure 1 The engrailed homeodomain protein-DNA complex (the template) as well as the modeled antennapedia and Mat α 2 protein-DNA complexes are shown as the center, the left, and the right molecules (respectively) in 1a. The different binding modes between the engrailed homeodomain and the DNA are shown in Fig. 1b. The interactions between ARG-54 of the Bicoid homeodomain and two different DNA binding sites (A1 and X1) are shown in Fig. 1c.



a



5(1)

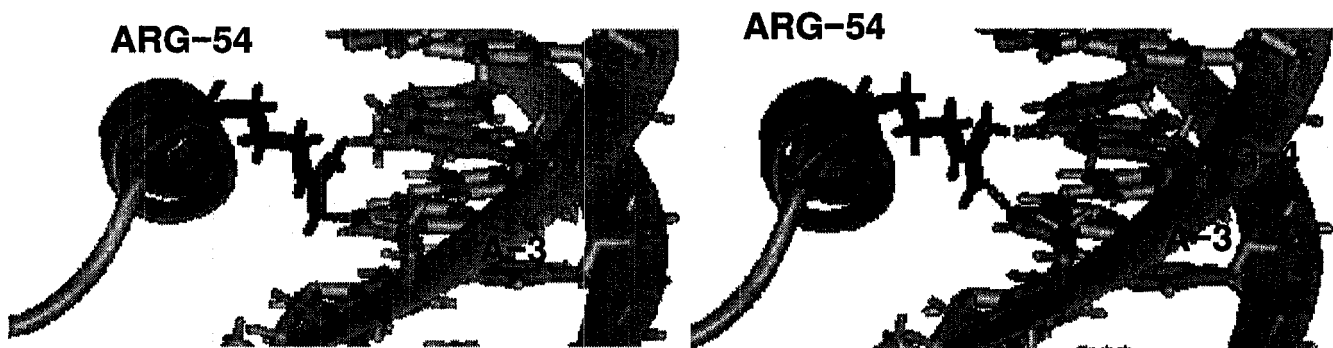
5(1)

7(3)

5(1)

4(0)

b



c

Figure 1